Journal of Digital Imaging (2021) 34:1183–1189
https://doi.org/10.1007/s10278-021-00465-y

Deep Learning Improves the Temporal Reproducibility of Aortic Measurement

Alex Bratt[1] · Daniel J. Blezek[1] · William J. Ryan[1] · Kenneth A. Philbrick[1] · Prabhakar Rajiah[1] · Yasmeen K. Tandon[1] · Lara A. Walkoff[1] · Jason C. Cai[1] · Emily N. Sheedy[1] · Panagiotis Korfiatis[1] · Eric E. Williamson[1] · Bradley J. Erickson[1] · Jeremy D. Collins[1]

Received: 29 November 2020 / Revised: 23 April 2021 / Accepted: 12 May 2021 / Published online: 28 May 2021

Abstract
Imaging-based measurements form the basis of surgical decision making in patients with aortic aneurysm. Unfortunately, manual measurement suffer from suboptimal temporal reproducibility, which can lead to delayed or unnecessary intervention. We tested the hypothesis that deep learning could improve upon the temporal reproducibility of CT angiography-derived thoracic aortic measurements in the setting of imperfect ground-truth training data. To this end, we trained a standard deep learning segmentation model from which measurements of aortic volume and diameter could be extracted. First, three blinded cardiothoracic radiologists visually confirmed non-inferiority of deep learning segmentation maps with respect to manual segmentation on a 50-patient hold-out test cohort, demonstrating a slight preference for the deep learning method ($p < 1e\text{-}5$). Next, reproducibility was assessed by evaluating measured change (coefficient of reproducibility and standard deviation) in volume and diameter values extracted from segmentation maps in patients for whom multiple scans were available and whose aortas had been deemed stable over time by visual assessment ($n = 57$ patients, 206 scans). Deep learning temporal reproducibility was superior for measures of both volume ($p < 0.008$) and diameter ($p < 1e\text{-}5$) and reproducibility metrics compared favorably with previously reported values of manual inter-rater variability. Our work motivates future efforts to apply deep learning to aortic evaluation.

Introduction

Thoracic aortic aneurysm is a morbid and costly disease requiring frequent surveillance due to the risk of dissection and rupture [1, 2]. Imaging surveillance, commonly with CT angiography (CTA), MR angiography, or echocardiography, is essential for risk stratification and pre-procedural planning. Two imaging parameters, maximum aneurysm diameter and diameter growth rate, are highly predictive of outcome and form the basis for surgical decision making [3]. Unfortunately, manual diameter measurement is challenging because of interobserver variability compounded by differences in measurement technique, which reduces temporal reproducibility and accounts for limited consistency even among experienced readers [4–8]. As a result, serial imaging follow-up may reveal spurious changes in aortic size.

✉ Alex Bratt
bratt.alexander@mayo.edu

[1] Department of Radiology, Mayo Clinic, 200 1stSt SW, Rochester, MN 55902, USA

Deep learning has shown promise in medical imaging as an adjunct to human perception and judgment. Deep learning systems trained on sufficiently large datasets can perform comparably to or superiorly to humans in some narrow radiology tasks [9, 10]. One of the key advantages of deep learning models is that they are deterministic, meaning the exact same input will always produce the exact same output, which makes aortic aneurysm surveillance an attractive target.

Unfortunately, deep learning is not a panacea, and at least two factors could make serial deep learning aortic measurement challenging: first, successive imaging exams on the same patient, despite in many cases being quite similar, are not exactly the same, owing to slight differences in positioning, scan parameters, and normal physiological changes. It is well-known that slight, even imperceptible, differences in input, such as those related to images acquired on different scanners, can lead to erratic model performance [11, 12]. This may negatively impact temporal reproducibility of automated measurements on serial imaging follow-up. Second, the performance of a deep learning model is only as good as its 'ground truth' training data. Thus, one might

expect a deep learning model to exhibit poor temporal reproducibility if that trait exists in its training set. This would be difficult to overcome since manual measurement is currently the gold standard method for quantification of aortic size. Though there is some data to suggest that deep learning is robust with respect to so-called 'noisy' or imperfect ground truth data [13], there is no guarantee that this will be the case for every clinical scenario to which it is applied.

In this retrospective study, we test the hypothesis that a deep learning model can improve upon the temporal reproducibility of serial manual aortic measurement using manual ground truth labels.

Materials and Methods

This research protocol was carried out under the supervision of our Institutional Review Board, which approved retrospective analysis of pre-existing datasets and waived the requirement for informed consent.

Training Data

We trained a standard deep learning segmentation model on 3051 retrospectively identified CTA volumes acquired at our institution between 2014 and 2020 ($n = 2835$ patients, ~1.08 million 2D images, mean age 62 years, 38% female). Each CTA was acquired on one of five Siemens (Forchheim, Germany) scanner models (Sensation 64, Edge Plus, Definition Edge, Definition Flash, Force) using standard technical parameters as per clinical protocol. Ground truth aortic segmentation maps were labeled by highly trained 3D lab technologists as part of the clinical workflow using commercially available post-processing software (TeraRecon, Durham, NC). Aortic labeling in our 3D lab is mostly manual but utilizes semi-automated tools such as region growing and interpolation. The only criteria for inclusion in the dataset were availability of clinical segmentation maps and anatomical coverage of the thorax. Segmentation maps included the aortic wall as well as the lumen. Prior to training, data was segregated based on medical record number to ensure there was no overlap between patients in the training cohort and either test cohort. No parameter tuning was applied after training nor was the model modified in any way following segmentation of either test cohort.

Qualitative Evaluation

The qualitative performance of the model was tested using a 50-scan hold-out test set (qualitative evaluation cohort, mean age 61 years, 42% female) for which three fellowship-trained cardiothoracic radiologists (LAW, YKT, PR) compared the visual quality of manual and automated segmentation maps. Manual and automated segmentation maps were randomly labeled either 'segmentation A' or 'segmentation B' and raters were blinded to the identity of each. Raters scored cases on an integer scale between −2 and 2 such that a score of 2 denoted that 'segmentation A' was highly superior and −2 highly inferior, with a score of zero denoting no difference. During analysis, labels were reassigned such that a score of 2 denoted that the automated segmentation was highly superior and −2 highly inferior.

Temporal Reproducibility

To evaluate the temporal reproducibility of measurements derived from segmentation maps, we identified a second cohort consisting of patients for whom multiple scans were available ('reproducibility cohort', $n = 57$ patients, 206 scans, mean age 65 years, 39% female). The mean number of scans per patient in the reproducibility cohort was 3.6 (min 2, max 8). The mean time interval between the first and last scan was 2.5 years (min 69 days, max 5 years). Additional inclusion criteria for patients in the reproducibility cohort were availability of two or more thoracic CTA scans and stability of thoracic aortic size, the latter of which was determined visually by a cardiothoracic radiologist (AB). We extracted from each segmentation map the volume of the ascending aorta (i.e., between the sinotubular junction and brachiocephalic artery origin) and the maximum double-oblique diameter of the distal ascending aorta just proximal to the brachiocephalic artery origin (Fig. 1).

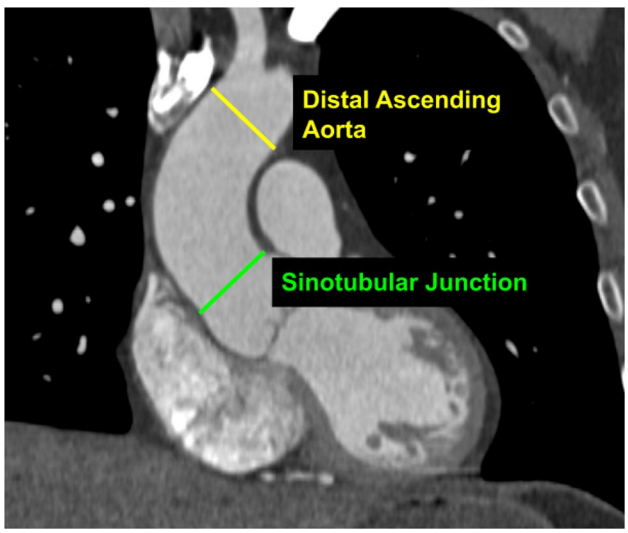

Fig. 1 Coronal CTA slice showing positions of sinotubular junction and distal ascending aorta

Deep Learning Model

The 2D deep learning segmentation model consisted of a U-net [14] backbone with an EfficientNet-B3 [15] encoder of depth seven. The model was built in PyTorch [16] using a publicly available segmentation library [17]. The input to the model was a single 512 × 512 CT slice and the output a binary segmentation map. During training, aggressive data augmentation was applied at runtime in the form of random crop (480 × 480), zoom, rotate, flip, Hounsfield unit window dilation/shift, and addition of Gaussian noise. Pixel values were rescaled to the interval between zero and one. A weighted cross-entropy function was used to compute the loss and weighting was empirically chosen based on prior experience (non-aorta weight = 0.3, aorta weight = 0.7). Incremental parameter updates were applied with RMSProp. The model was trained for 10 epochs, where epoch here denotes the un-augmented size of the training dataset. Training hardware included a 24 GB graphics card and a 24-core CPU.

Statistical Analysis

Comparisons between groups were made by using the Mann–Whitney rank test and signed-rank test for unpaired and paired data, respectively, and a two-sided p value threshold of 0.05 was chosen to indicate statistical significance. The coefficient of variation (i.e., the standard deviation divided by the mean) and standard deviation were used to compare reproducibility of manual and automated measurements per patient over longitudinal sets of scans. Any difference between measurements at different time points was attributed to user/model error. Reproducibility between successive pairs of measurements was assessed using the method of Bland and Altman [18], which yielded a mean difference and limits of agreement (LoA; mean ± 1.96 standard deviation). Dice score was used to compare overlap of manual and automated ascending aortic segmentation maps; other aortic segments were not compared in this way because of substantial heterogeneity in manual segmentation of non-aortic structures such as the kidneys and branch vessels. Cohen's kappa was used to evaluate agreement between raters. Statistical calculations were performed with the Python packages NumPy [19] and SciPy [20]. All CT scans were acquired in DICOM format and converted to NIfTI [21] for analysis.

Results

Training the model took approximately 19 days. Inference took approximately 10–30 s per case depending on anatomic coverage and slice thickness (versus 20–30 min for manual segmentation). Segmentation was successful in all scans in both cohorts, including in a wide array of pathologic states and anatomic variants such as right aortic arch, endovascular and open aortic repair, aneurysm, and severe atherosclerosis (Table 1, Fig. 2). This was confirmed by visual quality assessment; mean visual score across all 50 scans and three raters in the visual quality cohort was 0.28, representing a small but statistically significant preference for the automated images ($p < 1e\text{-}5$) with respect to the null hypothesis of zero mean score. Agreement between raters was substantial when defined as a difference in rating no more than two points (kappa = 0.79 [raters 1–2], 0.81 [raters 1–3], 0.78 [raters 2–3]), fair to moderate when defined as a difference of no more than one point (kappa = 0.44, 0.38, 0.42), and slight when considering any difference as disagreement (kappa = 0.13, 0.09, 0.08). Mean dice score between manual and automated ascending aortic segmentation maps was 0.96.

Automated reproducibility was superior to manual reproducibility. In the reproducibility cohort, mean coefficient of variation (and standard deviation) for ascending aortic volume was 6.0% (4.3 mL) for manual vs. 4.5% (3.3 mL) for automated ($p = 0.008$). Mean coefficient of variation (and standard deviation) for distal ascending aortic diameter was 4.2% (1.7 mm) for manual vs. 2.1% (0.8 mm) for automated ($p < 1e\text{-}5$). There were small but statistically significant differences between paired manual and automated measurements (mean diameter = 39mm [automated] vs. 40mm [manual], $p = 0.02$; mean volume = 79 mL [automated] vs 78 mL [manual], $p < 1e\text{-}7$).

Bland–Altman analysis also demonstrated superior agreement for the deep learning measurements (Fig. 3). Limits of agreement were ± 6.6 mm for manual diameter measurement, ± 3.1 mm for automated diameter measurement, ± 14.5 mL for manual volume measurement, and ± 11.6 mL for automated volume measurement. Change between successive measurements was not significantly different from zero ($p > 0.05$ for all permutations of [automated,

Table 1 Prevalence of aortic findings in the qualitative evaluation cohort

Finding	Number of patients (%)
Normal aorta	11 (22)
Aortic aneurysm	20 (40)
Open aortic repair	17 (34)
Aortic dissection	9 (18)
Coarctation repair	2 (4)
Coral reef calcification	1 (2)
Right aortic arch	1 (2)
Aortic endograft	1 (2)
Aortitis	1 (2)
Aortic occlusion	1 (2)

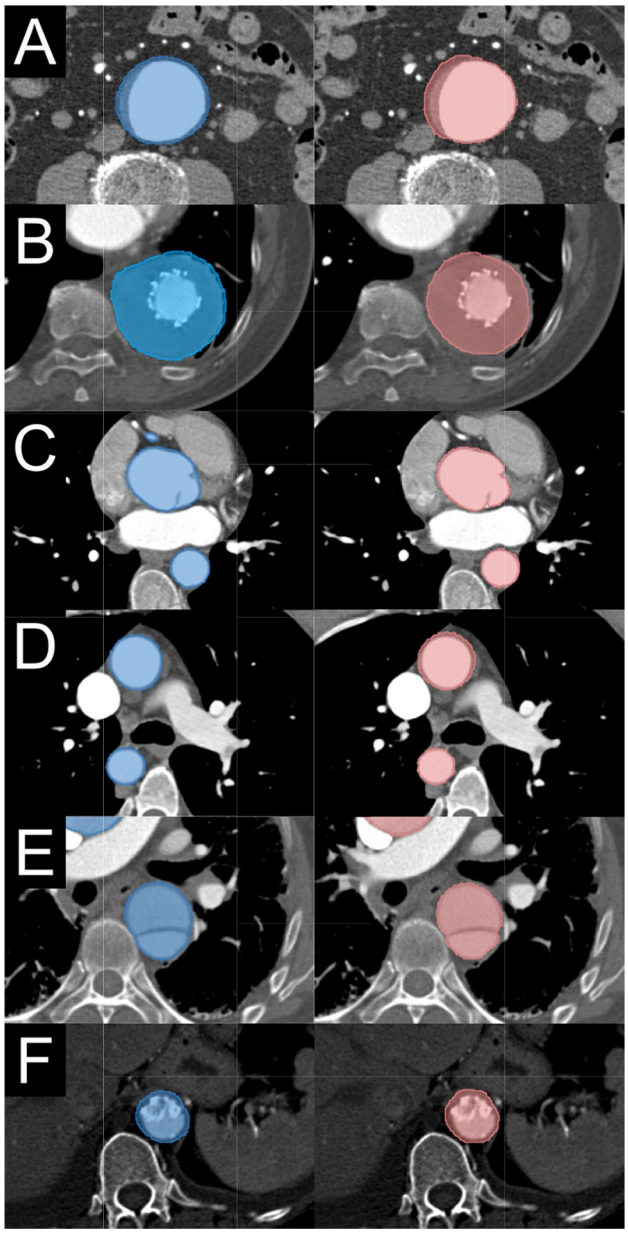

Fig. 2 Representative examples of automated (blue, left column) vs. manual (red, right column) segmentations. Note good agreement in a wide spectrum of anatomy and pathology, including aneurysm **A**, endograft **B**, complex 3D structure (**C**, aortic root), right aortic arch **D**, dissection **E**, and bulky luminal calcifications **F**

manual] × [diameter, volume]), confirming aortic size stability.

Discussion

This is the first study to analyze temporal reproducibility of deep learning for aortic measurement. Our results support the use of deep learning for aortic quantification. Deep learning segmentation quality was evaluated in three ways. First, three raters visually compared automated and manual segmentations in a 50-scan cohort, demonstrating non-inferiority with respect to manual segmentation. This provided a favorable assessment of model performance over a wide spectrum of anatomy and pathology.

Third, we compared temporal reproducibility of automated- and manual-derived measurements in a cohort of patients with stable aortic size over multiple scans. We showed that measurement reproducibility was superior for measurements derived from automated segmentations and that automated reproducibility compared favorably to values reported in the literature. Quint et al. [4] showed that intra-observer standard deviation ranged between 1.2 and 2.7 mm for diameter measurement at several proximal thoracic aortic landmarks, comparing well with the standard deviation of 0.8 mm our model achieved. Mora et al. [7] reported intra-observer LoA between ± 1.99 and ± 4.84 mm for double oblique linear measurement of abdominal aortic aneurysms, which is similar to the LoA ± 3.1 mm we found for successive automated diameter measurements. Parr et al. [8] examined the reproducibility of abdominal aortic aneurysm volume measurement and found an intra-rater coefficient of variation of 2.8%, which is slightly lower than the 4.5% mean coefficient of variation we obtained for automated ascending aortic volume measurement. We attribute our higher coefficient of variation to (a) the fact that we compare volumes across different exams instead of across different observations of the same exam, (b) the high sensitivity of volume measurement to slight differences in aortic size that occur between different exams, and (c) the greater influence of the cardiac cycle on proximal thoracic aortic volume as opposed to abdominal aortic volume.

Distal aortic diameter and ascending aortic volume were chosen as targets of analysis for four reasons: first, some of the scans in our dataset were ungated and the ascending aorta is not as susceptible to motion artifact as more proximal landmarks such as the aortic annulus or root. Second, the distal ascending aorta is a standardized and clinically useful measurement location [22, 23]. Third, the distal ascending aorta is defined by a precise and easily identifiable landmark (i.e., the brachiocephalic artery origin), unlike some other clinically useful landmarks like the mid ascending or descending aorta, which reduced variability related to manual plane selection and allowed the attribution of any temporal change to model error rather than plane selection. Fourth, in our dataset, ascending aortic size tended to be more stable over time than segments of the aortic arch and descending thoracic aorta, which enabled the exclusion of fewer scans. Ground truth segmentations considered the vessel wall as part of the aorta, as is standard in patients with aneurysmal disease [23–25], and this was successfully reproduced by the deep learning model and reflected in all measurements (Fig. 2).

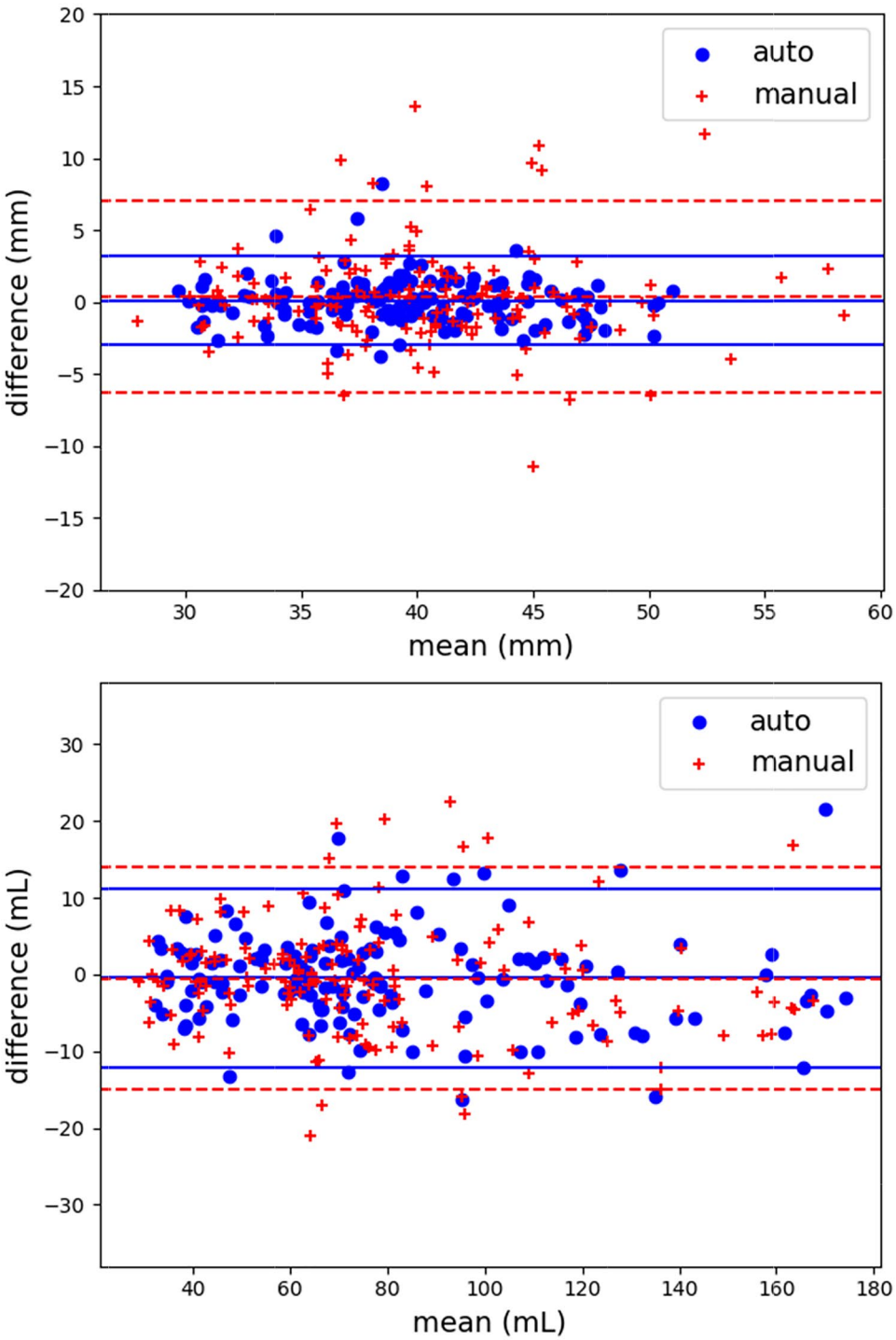

Fig. 3 Results of Bland–Altman analysis. The y-axis represents difference between paired successive measurements and the x-axis represents the mean of each pair. Middle and flanking lines represent the mean and the mean ± 1.96 * SD, respectively (solid for auto and dashed for manual). Top panel is diameter and bottom is volume. Limits of agreement for diameter are ± 3.1 mm for automated and ± 6.6 mm for manual. Limits of agreement for volume are ± 11.6 mL for automated and ± 14.5 mL for manual. Note the wider spread of manual differences (crosses) relative to automated (dots), reflecting poorer manual reproducibility

The above findings were non-obvious; while the efficiency and time-savings of deep learning are well known, the consistency and reproducibility of these models remain unproven, largely due to limitations such as brittleness and poor generalizability as well as limitations related to training data quality. We confirm deep learning's impressive robustness to label noise in the setting of aortic size measurement, showing improved reproducibility compared to imperfectly annotated training data. We attribute the success of our model at least partially to the large size of the training dataset, which enabled a dense sampling of the patient population and prevented overfitting.

Our work motivates future efforts to use deep learning to perform automated aortic measurement. We hope that future fully automated models will enable wider adoption and study of volumetric analysis, which is known to be

more sensitive to aortic size change than linear measurement [26–28] but poses practical challenges because of the necessity of time-consuming manual segmentation. We also hope that automation will usher in more sophisticated normalization methodologies including internal indices such as aortic length and vertebral body area/volume, as normal ranges for absolute aortic measurements fluctuate considerably with body size [29].

Limitations are as follows: first, while our model performs aortic segmentation in a fully automated way, manual plane selection is necessary to extract measurements. Future work will focus on automating the remaining steps in this process. Second, we make the imperfect assumption that any change in aortic size over time can be attributed to measurement error. While we made every effort to exclude patients with growing aortas, some imperceptible changes may have been missed and this could affect our results. Thankfully, statistical analysis confirms that small differences between successive aortic measurements are likely insignificant. Third, we do not report data on external validation. This study focuses on demonstrating a general property of deep learning applied to aortic measurement using commonly available software and hardware. Thus, these results are more likely to be reproducible in an external setting than the performance of a specific deep learning model. Fourth, our training data consists of segmentation maps performed by technologists for the purpose of qualitative pre-procedural evaluation and not size measurement. Paradoxically, this may serve as a benefit as well as a limitation since it likely increased ground truth noise and thereby showed more clearly the robustness of our deep learning model with respect to training data imperfections. Thankfully, the performance of the model does not seem to suffer as measurements extracted from automated segmentations compare well to reproducibility metrics reported in the literature. Fifth, we did not assess aortic diameter at all clinically relevant and canonical aortic stations, such as the ascending aorta at the level of the pulmonary arteries. We acknowledge that comparisons to prior work are problematic given variable technique and anatomic focus (abdominal aorta vs. thoracic aorta) but we believe such comparisons are informative nonetheless.

Conclusions

We have shown that a deep learning aortic segmentation model improves upon the visual quality and temporal reproducibility of its training data and compares favorably with manual reproducibility data reported in the literature.

Acknowledgements We would like to thank Mayo Clinic for funding this study.

Author Contribution Guarantor of integrity of entire study: A.B.; study design: all authors; data curation: A.B., D.J.B., W.J.R., K.A.P.; image interpretation: A.B., P.R., Y.K.T., L.A.W.; statistical analysis: A.B.; manuscript drafting/editing: all authors.

Funding A.B received institutional support from Mayo Clinic.

Data Availability Anonymized data can be made available upon request.

Code Availability Code pertaining to the deep learning model is freely available online as described in the body of the manuscript. Scripts for data acquisition, preprocessing, and postprocessing can be made available upon request.

Declarations

Ethics Approval This research protocol was performed under the supervision of our institutional review board, which approved the retrospective analysis of pre-existing data and waived the requirement for informed consent.

Conflict of Interest PR: royalties from Elsevier for book and journal editing (not related to present manuscript). JDC: co-investigator for an institutional grant from Siemens Healthineers; stock owner in Ceta, a non-publicly traded healthcare I company; travel costs from Siemens Healthineers to attend a CT research summit in Germany in October 2019 (none related to present manuscript).

References

1. Abdulameer H, Taii HA, Al-Kindi SG, Milner R. Epidemiology of fatal ruptured aortic aneurysms in the United States (1999-2016). Journal of Vascular Surgery. 2019 Feb 1;69(2):378-384.e2.
2. LeMaire SA, Russell L. Epidemiology of thoracic aortic dissection. Nature Reviews Cardiology. 2011 Feb;8(2):103–13.
3. Aggarwal S, Qamar A, Sharma V, Sharma A. Abdominal aortic aneurysm: a comprehensive review. Exp Clin Cardiol. 2011;16(1):11–5.
4. Quint LE, Liu PS, Booher AM, Watcharotone K, Myles JD. Proximal thoracic aortic diameter measurements at CT: repeatability and reproducibility according to measurement method. Int J Cardiovasc Imaging. 2013 Feb;29(2):479–88.
5. Dugas A, Therasse É, Kauffmann C, Tang A, Elkouri S, Nozza A, et al. Reproducibility of abdominal aortic aneurysm diameter measurement and growth evaluation on axial and multiplanar computed tomography reformations. Cardiovascular and interventional radiology. 2012;35(4):779–87.
6. Shimada I, Rooney SJ, Farneti PA, Riley P, Guest P, Davies P, et al. Reproducibility of thoracic aortic diameter measurement using computed tomographic scans. Eur J Cardiothorac Surg. 1999 Jul 1;16(1):59–62.
7. Mora C, Marcus C, Barbe C, Ecarnot F, Long A. Measurement of maximum diameter of native abdominal aortic aneurysm by angio-CT: reproducibility is better with the semi-automated method. European Journal of Vascular and Endovascular Surgery. 2014 Feb 1;47(2):139–50.
8. Parr A, Jayaratne C, Buttner P, Golledge J. Comparison of volume and diameter measurement in assessing small abdominal aortic aneurysm expansion examined using computed tomographic angiography. European Journal of Radiology. 2011 Jul 1;79(1):42–7.

9. Ardila D, Kiraly AP, Bharadwaj S, Choi B, Reicher JJ, Peng L, et al. End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography. Nature Medicine. 2019 May 20;1.
10. Kuo W, Häne C, Mukherjee P, Malik J, Yuh EL. Expert-level detection of acute intracranial hemorrhage on head computed tomography using deep learning. PNAS. 2019 Nov 5;116(45):22737–45.
11. Zech JR, Badgeley MA, Liu M, Costa AB, Titano JJ, Oermann EK. Variable generalization performance of a deep learning model to detect pneumonia in chest radiographs: a cross-sectional study. PLOS Medicine. 2018 Nov 6;15(11):e1002683.
12. Yan W, Huang L, Xia L, Gu S, Yan F, Wang Y, et al. MRI manufacturer shift and adaptation: increasing the generalizability of deep learning segmentation for MR images acquired with different scanners. Radiology: Artificial Intelligence. 2020 Jul 1;2(4):e190195.
13. Jiang L, Di Huang ML, Yang W. Beyond synthetic noise: deep learning on controlled noisy labels. In ICML; 2020.
14. Ronneberger O, Fischer P, Brox T. U-Net: convolutional networks for biomedical image segmentation. In: Navab N, Hornegger J, Wells WM, Frangi AF, editors. Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. Springer International Publishing; 2015. p. 234–41. (Lecture Notes in Computer Science).
15. Tan M, Le Q. {E}fficient{N}et: rethinking model scaling for convolutional neural networks. In: Chaudhuri K, Salakhutdinov R, editors. Proceedings of the 36th International Conference on Machine Learning [Internet]. Long Beach, California, USA: PMLR; 2019. p. 6105–14. (Proceedings of Machine Learning Research; vol. 97). Available from: http://proceedings.mlr.press/v97/tan19a.html
16. Paszke A, Gross S, Chintala S, Chanan G, Yang E, DeVito Z, et al. Automatic differentiation in PyTorch. In: NeurIPS Autodiff Workshop. 2017.
17. Yakubovskiy P. qubvel/efficientnet [Internet]. 2020 [cited 2020 Sep 30]. Available from: https://github.com/qubvel/efficientnet
18. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet. 1986 Feb 8;1(8476):307–10.
19. Harris CR, Millman KJ, van der Walt SJ, Gommers R, Virtanen P, Cournapeau D, et al. Array programming with NumPy. Nature. 2020;585:357–62.
20. Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D, et al. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. Nature Methods. 2020;17:261–72.
21. Cox R, Ashburner J, Breman H, Fissell K, Haselgrove C, Holmes C, et al. A (Sort of) new image data format standard: NIfTI-1: WE 150. Neuroimage. 2004;22.
22. Hiratzka Loren F., Bakris George L., Beckman Joshua A., Bersin Robert M., Carr Vincent F., Casey Donald E., et al. 2010 ACCF/AHA/AATS/ACR/ASA/SCA/SCAI/SIR/STS/SVM Guidelines for the diagnosis and management of patients with thoracic aortic disease. Circulation. 2010 Apr 6;121(13):e266–369.
23. Hout MJ van, Scholte AJ, Juffermans JF, Westenberg JJ, Zhong L, Zhou X, et al. How to measure the aorta using MRI: a practical guide. Journal of Magnetic Resonance Imaging. 2020;52(4):971–7.
24. Freeman LA, Young PM, Foley TA, Williamson EE, Bruce CJ, Greason KL. CT and MRI assessment of the aortic root and ascending aorta. American Journal of Roentgenology. 2013 May 23;200(6):W581–92.
25. Díaz-Peláez E, Barreiro-Pérez M, Martín-García A, Sanchez PL. Measuring the aorta in the era of multimodality imaging: still to be agreed. J Thorac Dis. 2017 May;9(Suppl 6):S445–7.
26. Kitagawa A, Mastracci TM, Allmen R von, Powell JT. The role of diameter versus volume as the best prognostic measurement of abdominal aortic aneurysms. Journal of Vascular Surgery. 2013 Jul 1;58(1):258–65.
27. Geisbüsch S, Stefanovic A, Schray D, Oyfe I, Lin H-M, Luozzo GD, et al. A prospective study of growth and rupture risk of small-to-moderate size ascending aortic aneurysms. The Journal of Thoracic and Cardiovascular Surgery. 2014 Jan 1;147(1):68–74.
28. Trinh B, Dubin I, Rahman O, Ferreira Botelho MP, Naro N, Carr JC, et al. Aortic volumetry at contrast-enhanced MR angiography: feasibility as a sensitive method for monitoring bicuspid aortic valve aortopathy. Invest Radiol. 2017 Apr;52(4):216–22.
29. Zafar MA, Li Y, Rizzo JA, Charilaou P, Saeyeldin A, Velasquez CA, et al. Height alone, rather than body surface area, suffices for risk estimation in ascending aortic aneurysm. The Journal of Thoracic and Cardiovascular Surgery. 2018 May 1;155(5):1938–50.